INTEGRATING SOCIAL MEDIA AND RAINFALL DATA TO UNDERSTAND THE IMPACTS OF SEVERE WEATHER IN ARGENTINA

BY

STELLA LINA CHOI

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Atmospheric Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Adviser:

Associate Professor Stephen W. Nesbitt

ABSTRACT

Subtropical South America experiences some of the most intense deep convection in the world. In terms of socio-economic impacts, flooding was the most destructive natural disaster in Argentina between 1980-2010; it affected 343 million people, and caused over $222 billion USD in damages. Furthermore, the national weather service in Argentina has a history of operational issues, the most glaring of which is that a single office in Buenos Aires is responsible for forecasting for the entire country. Consequently, there is a large disconnect between the weather service and the public, which impedes effective severe weather communication. We leverage social media data to understand the public's perception and Twitter activity during heavy rainfall events. Previous studies have investigated the role of social media in circulating critical information during emergencies; however, few have looked into the impact beyond cities in the United States. Rainfall and Twitter activity demonstrate a direct relationship, yet complex, non-linear interactions are likely impacting the results. A new metric, tweeting, efficiency, is developed to account for the inherent lull in social media activity during hours people are most likely asleep. Geo-tagged posts also provide supplemental information, which is particularly advantageous for this region, as it suffers from sparse and biased data.

ACKNOWLEDGEMENT

I would like to thank my adviser, Dr. Steve Nesbitt, for his guidance with this research project, and for the invaluable opportunities during graduate school. Discussions with Dr. Paola Salio, Marcos Saucedo, Julia Chasco, Ignacio Gatti, and forecasters at the Servicio Meteorológico Nacional Argentina have been immensely helpful in informing the direction of this study. I would also like to thank family, friends, and former and current mentors for their encouragement and support throughout this process. A National Science Foundation Graduate Research Fellowship, and the Tinker Foundation's Field Research Grant supported this research.

TABLE OF CONTENTS

1. INTRODUCTION

Heavy rainfall across Buenos Aires, Argentina caused record-breaking flash floods that made international headlines on 2-3 April 2013. This *unprecedented catastrophe* inflicted overwhelming damage on the provincial capital, La Plata, which has a population of more than one million. Reported floodwater levels reached about 2 meters, and the intense rainfall covered over 60 km. La Plata was inundated by *over four times its record April rainfall within a few* hours [*Gilbert*, 2013]. The societal impacts of the April 2013 La Plata flood were devastating; however, subtropical South America is known to experience some of the most intense deep convection in the world [*Zipser et al.*, 2006; *Rasmussen et al.*, 2014].

Mesoscale convective systems (MCSs) in the United States are widely recognized for being responsible for heavy rainfall and flash floods [*Houze et al.*, 1990]. However, there has been minimal progress in understanding the mechanisms responsible for the hydrologic impacts of Southeastern South America (SESA) MCSs, let alone their predictability [*Rasmussen and Houze*, 2011; *Rasmussen et al.*, 2014]. SESA MCSs are more persistent and 60% larger than those over the US [*Velasco and Fritsch*, 1987]. These storms typically form during strong low-level heat and moisture advection from the Amazon near and apart from the Andes mountain range, and propagate upstream backbuilding against the South American low-level jet (SALLJ) [*Anabor et al.*, 2007]. When the SALLJ is active, precipitation over SESA is enhanced [*Nicolini and Saulo*, 2006], MCSs are larger, and persist longer, and there is a maximum in convective frequency over northeastern Argentina [*Salio et al.*, 2007].

Subtropical South America remains vulnerable to flash flooding in part due to its complex and enigmatic meteorology, and highly urbanized populace. According to statistics provided by *Prevention Web*, flooding was the most destructive natural disaster in Argentina between 1980-2010. It affected 343 million people and incurred over $222 billion USD in damages. Furthermore, biased satellite measurements and sparse data render forecasting a challenge in Argentina. Developers of the Tropical Rainfall Measuring Mission Precipitation Radar (TRMM PR) algorithm recognize that rain rates are underestimated, especially over land [*Iguchi et al.*, 2009]. Previously studies have quantified a nearly 40% bias in deep convective rain rates in subtropical South America [*Rasmussen et al.*, 2013].

Compounding to SESA's vulnerability to flooding and the inherent inaccuracies in satellite measured precipitation, the Argentine national weather service, Servicio Meteorológico Nacional (SMN), faces challenges stemming from a series of overarching issues. The SMN suffers from a history of operational problems, the most glaring of which is that a single office in Buenos Aires is responsible for forecasting for the entire country. According to one SMN meteorologist, "[In Argentina] it is so challenging because of the large area we need to forecast for. We are always going to miss something." The lack of manpower in forecasting further amplifies other obstacles including outdated infrastructure and services, and interactions with the media and the public.

Argentine forecasters say the media reaches a wider audience. For example, as of April 2016 one media meteorologist has nearly 80,000 more Twitter followers than the SMN. One SMN employee claims, "People pay much attention to TV forecasts. I stress upon this because it is what affects us the most. The language is fundamental." This relationship becomes problematic

when sensationalism follows, as "mainly chaos is generated when the media alerts the public." Furthermore, "the public is generally not aware that warnings are issued because they do not regularly access the [weather service] webpage." Considering a successful forecast not only involves accurate science, but an appropriate response by the public, it is imperative that Argentines' concerns regarding severe weather are heard.

Social media adds a new dimension to human interactions, and these data can be leveraged to glean insight into the Argentine public's perception surrounding severe weather. Twitter is one microblogging platform that launched in 2006 where users post concise messages that do not exceed 140 characters, or so-called *tweets* [*Guan and Chen*, 2014]. Over 320 million people were monthly active users as of December 2015 [*Twitter website*, 2016]. The Twitter archive grows by over half a billion everyday through the Twitter website or an application [*Library of Congress*, 2013]. Some of the earlier studies on Twitter focused on data mining and understanding the activity of its users [*Java et al.*, 2007]. Subsequent studies categorized users into discrete types based on their activity [*Krishnamurthy et al*., 2008]. Twitter users have been found to interact with only a small group of their friends on the site [*Huberman et al.*, 2008]. Content analyses demonstrated that Twitter is comparable to traditional news medium, however news surrounding unusual events spread more rapidly [*Zhao et al.*, 2011]. Researchers began extending studies into event-specific cases, e.g. crisis informatics, by coupling society and technology to domains including public health, natural disasters, politics, and economy [*Hagar*, 2009; *Kryvasheyeu et al*., 2016; *Bond et al*., 2012; *Conover et al*., 2013; *O'Connor et al*., 2010; *Llorente et al*., 2015].

Twitter activity surrounding mass convergence and emergency events develop into information diffusion, as there are less reply tweets to specific users, and a larger volume of URLs compared to general collection of tweets [*Hughes et al.*, 2009]. In general, URLs have been appearing more in tweets, which is evidence of users using the platform for more information sharing [Hughes et al., 2009]. A subset of this research focuses on the relationship between Twitter activity and natural disasters. A study on Hurricane Sandy observed a strong spatiotemporal correlation between social media activity and damage, especially near coasts and large cities [*Guan and Chen*, 2014]. During crises, the volume and velocity of tweets are exceptionally high, and it remains a challenge to filter unrelated messages efficiently through machine learning techniques [*Herfort et al.*, 2014]. Thus, studies have developed geographical methods to add extra constraints to optimize filtering during emergency situations. For example, a spatial distribution of tweets about floods was vastly different than unrelated messages [*Herfort et al.*, 2014]. The spatiotemporal lifecycle of Twitter activity surrounding Hurricane Sandy determined the microblogging network complements traditional media, but provides advantages for emergency managers due to its accessibility, and simplicity [*Kryvasheyeu et al*., 2016].

Previous works have provided insight into general Twitter use, interactions among users, and event-based analyses during extraordinary events. This study aims to listen to the Argentine public through large volume of Twitter data, and translate their concerns into actionable insight surrounding heavy rainfall and flooding. Big data and social media have the potential to add insight into understanding behavior; however, it is not a stand-alone solution. Analyses must be performed with caution due to continuously shifting social norms and technological updates [*Lazer et al*., 2014]. Understanding social interactions remains challenging due to the

nonlinearities and dependencies among variables and data [*Lazer et al.*, 2014]. As such, this study combines personal narratives of forecasters to maintain a more human and personal component to data analysis.

Previous study in Buenos Aires reveals that the general public understands forecasts in terms of impacts, as they call Argentine forecasters with practical questions. Additionally, past events such as the 2003 Santa Fe floods in Argentina have acted as catalysts to restructure local weather services and motivate preventative action [*Choi*, 2014]. Providing a voice to the public to minimize impacts of severe weather in Argentina is imperative to establish effective hydrologic disaster decision-making policies, and mitigate impacts in vulnerable regions lacking infrastructure and preparation, e.g. SESA.

We hypothesize: 1. footprints of severe weather events are evident in social media; 2. regional analysis of social media posts reveals insight and quantitative information that complements traditional weather observations (e.g., satellites, radars) cannot detect; 3. social media provides a venue to listen to the public's concerns and perception surrounding hazardous weather. This study focuses on severe rainfall and flooding events in nine provinces in central Argentina. Considering the ubiquity of social media, there is a growing body of literature leveraging these data to glean insight into behavior during natural hazards and emergencies. However, these studies tend to focus on cities in the United States during extraordinary events. This research aims to glean further insight into heavy precipitation events in subtropical South America, a region that suffers from sparse data and complex meteorology, by integrating provincial Twitter data, and rainfall measurements.

2. DATA AND METHODOLOGY

This study aims to determine to what extent public tweets can heighten situational awareness surrounding severe weather events in Argentina, particularly in data and media sparse regions outside of Buenos Aires province. To accomplish this goal, we combine macroscale data (tweets and rainfall measurements) and microscale data (personal narratives and quotes). Satellite derived rainfall estimates from NASA's Tropical Rainfall Measuring Mission (TRMM) 3B42 version 7 are averaged over a bounding box for each province [*Simpson et al.*, 1988]. Table 2 summarizes the latitude and longitude bounds defined for each province to retrieve rainfall rates. This study leverages Twitter for its efficiency over other social media platforms, as posts are limited to a strict character limit, yet are able to contain different mediums: text, photo, video, and URLs. Furthermore, studies have shown Twitter to be an effective platform for rapid communication during crises, especially near cities [*Guan and Chen*, 2014]; however, few have given attention to communities and severe events that may have not received as much media attention. In Argentina, the single office in Buenos Aires tends to bias forecasts and warnings heavily towards the capital [*personal communication with Marcos Saucedo (SMN)*, 2015]. In addition to potential real-time applications, storm reports can also aid Buenos Aires forecasters by providing additional spatiotemporal information of the impacts of severe weather, and to evaluate and verify forecasts and warnings.

Crimson Hexagon is a data provider that stores a full archive of Tweets from December 2013, also known as a Twitter Firehose. The Twitter Streaming Application Programming Interface (API) allows the public to collect incoming tweets in near real-time. There are advantages and disadvantages associated with both methods. The Twitter Streaming API is free to the public,

however only returns 1% of all tweets at a given time, though the algorithm Twitter employs to sample remains opaque [*Morstatter et al*., 2013]. Alternatively, the Twitter Firehose provides a complete dataset of public tweets, but at a steep monetary cost [*Morstatter et al*., 2013]. Furthermore, geotagged tweets accounts for a small fraction of both Streaming and Firehose datasets, 3.17% and 1.45%, respectively [*Leetaru et al.*, 2013; *Morstatter et al*., 2013]. Crimson Hexagon extracts at most 10,000 tweets per bulk export, and geolocation is only on the provincial level, which lacks fine spatial resolution. Regardless, this study opts for Twitter Firehose data for archived information for a more complete story surrounding severe weather in Argentina.

We extract Twitter data from Crimson Hexagon from December 2013 to December 2015 by querying specific keywords in Spanish focusing on extreme precipitation and flooding in nine provinces. Figure 1 shows the nine provinces, which include: the Autonomous City of Buenos Aires (CABA), Buenos Aires, Cordoba, Corrientes, Entre Rios, La Pampa, Mendoza, San Luis, and Santa Fe. The comprehensive list of keywords and phrases is as follows: *precipitación, precipitacion, precipitaciones, lluvia, lluvias, tormenta, tormentas, granizo, inundación, inundacion, inundaciones, tormentas severas, tormenta severa, tormenta intensa, tormentas intensas, tormentas fuertes, intensas lluvias, alerta meteorológica, and alerta meteorologica*. Non-accented words were included to account for potential spelling errors, as correspondence on the platform tend to be casual. To minimize the retrieval of unrelated tweets, words such as intense and strong are included in phrases, e.g. “intense storms”.

As discussed earlier, the Twitter Firehose is not a stand-alone solution to retrieving a comprehensive collection of tweets. Crimson Hexagon allows up to 10,000 tweets pre bulk export; thus, provinces with high population density suffer from lower ratios of Analyzed Tweets to Total Filtered Tweets that were found (table 1). The ratio also likely depends on the keywords and the number of keywords. The ratio would decrease if more keywords were queried, as the Total Filtered Tweet count would increase with more tweets meeting the criteria. Furthermore, the sample size Analyzed Tweets was reduced, as all tweets by bots (defined as unique authors with over 50 tweets in one month by province). The percentages of automated tweets written by bots omitted varied by province, and strongly depended on population density, which is likely due to the arbitrary number set to detect bots. The Autonomous City of Buenos Aires had the lowest percentage of tweets written by bots (6.95%), while Mendoza had the largest (26.79%). These percentages were calculated as the ratio of deleted bot tweets to the total number of tweets originally found for analysis.

3. RESULTS AND DISCUSSION

3.1 Regional analysis of word frequency

A heatmap of term frequency by province is summarized in figure 2. The frequencies are calculated over the two years of data as a ratio between keyword count and total word count of all tweets by province. The frequencies of the following terms are plotted: *rain, storm, flood, alert, meteorology, hail, intense, wind, and sleep*. The terms *wind* and *sleep* were not explicitly queried when collecting tweets, but were byproducts in our data collection, and more ubiquitous than terms that were included in the keyword search. For example *sleep* is more common among tweets than *meteorology*, which was specifically searched.

The keyword *rain* is the most ubiquitous across all provinces with a maximum in Entre Rios, and a minimum in the Autonomous City of Buenos Aires (CABA). This range could be influenced by the differences in volume of tweets and population density between the two provinces. Tweets related to *flooding* were the most ubiquitous in CABA. The dense population, urban geography, and localized deep convection in the region renders CABA the most vulnerable to flash floods. Furthermore, media sensationalism, and biased forecasts and alerts towards the capital city could be amplifying these factors. *Rasmussen et al*. [2014] uncovered a large volume of media reports surrounding floods in La Pampa, but the frequency *flood* in tweets is relatively low. Unlike the flash floods in urban cities, e.g. Buenos Aires, La Pampa experiences slow-rise floods due to its agricultural terrain [*Latrubesse and Brea*, 2009]. The rainfall over this region is also contributed by broad stratiform and wide convective cores [*Romatschke and Houze*, 2010; *Rasmussen and Houze*, 2011]. As such, the intensity and temporal differences between the two types of flooding could be influencing the dialogue in the media and Twitter differently. The

impacts of flash flooding in Buenos Aires could be more visible to more people, which might promote discussions and sharing information, e.g. photos and alerts, on Twitter. Watching intense flash floods most likely influences people's emotions as well, which could modulate Twitter activity. On the other hand, media reports from La Pampa of slow-rise floods could be analyzed to determine if the reports are mostly related to the impacts on the province and agriculture.

Mendoza is located at the foothills of the Andes Mountains, and is a region vulnerable to hailstorms. Studies have also found a maximum in media reports of hail in this province [*Rasmussen et al*., 2014]. Similar to news reports, the maximum frequency of the term *hail* in tweets is also in Mendoza. The frequency of *hail* appearing in tweets does not appear to coincide with *alert*. The keyword *alert* is most tweeted about in Cordoba. This province has the second highest population density in Argentina; however, the weather service responsible for forecasts and alerts is in Buenos Aires. As such, severe weather alerts in Cordoba could be disseminated more through Twitter. The keyword *meteorology* was also the most ubiquitous in Cordoba.

As discussed earlier two terms, *sleep* and *wind*, appeared frequently in tweets, but were not explicitly queried in our data collection. *Wind* appeared most frequently in tweets from La Pampa. *Rasmussen et al.* [2014] showed that this region is the South American equivalent of "tornado alley". As such, tweets about winds could correspond with the organized MCSs that form in this region that is susceptible straight-line wind damage and tornadoes. Tweets in Buenos Aires province had the highest frequency of the term *sleep*. Rain rate maximum in Buenos Aires province is around 3:00 local standard time (LST) consistently throughout

summer, fall, and spring (figure 3). Thus, it may be that Argentines are being disrupted in their sleep during these rainfall events.

3.2 Relationship between rainfall and tweets

We analyze the relationship between rainfall and Twitter activity for Buenos Aires province by season: DJF (summer), MAM (fall), JJA (winter), and SON (spring). Figure 3 shows an hourly time series of the total rain rates (mm/hr) and tweet count. The total tweet count is calculated for every three hours, i.e. the total tweet count at 3:00 (LST) accounts for all tweets between 3:00 to 6:00 (non-inclusive). Each tweet is further categorized into sentiments, which are determined by Crimson Hexagon as positive, neutral, and negative.

Peak rain rates generally occur in the early morning throughout all seasons, except the spring from September to November, when peak rainfall is in the evening. Throughout all seasons and hours of the day, neutral tweets account for the highest percentage. There is a slight increase in negative tweets in the early morning hours, which could relate to the maximum frequency of *sleep* appearing in the tweets in Buenos Aires province. Furthermore, studies have shown that generally, people's moods tend to worsen as the day progresses [*Golder*, 2011]. These are two potential influences affecting tweeting behavior, although there are likely much more complex nonlinear interactions impacting these results.

Rolling averages were calculated for tweet counts and rain rates for $\pm$ 3 hours to minimize the noise present in the rainfall and Twitter data. This noise manifests itself as time offsets between the maximum rain rates and Twitter activity, as it depends whether people are tweeting more

before, during, or after events. Higher Twitter activity before events could reflect more media attention and sensationalism indicating extreme events that are approaching. This trend could also be more prominent near Buenos Aires, and densely populated urban areas. Maximum Twitter activity coinciding with the heaviest rainfall could be evidence of flash floods, or shorter-scale deep convective events that are intense, but difficult to forecast in advance. Events during which Twitter activity lags peak rain rates could indicate broad stratiform precipitation, which is relatively less intense but contributes higher volume of rainfall [*Romatschke and Houze*, 2010; *Rasmussen and Houze*, 2011]. Relating peaks in rainfall with tweeting activity reduced this noise (figure 4), and coupled the peaks as single events as rolling averages within $\pm$3 hour window of time (figure 5). Figure 4 shows two scatter plots relating average rain rates as a function of tweet count for the raw data (left), and rolling means (right) by hour of day in local standard time. R-squared values of the rolling averages are relatively low, and no particular relationship could be detected. Further analyses by province related to population density and type of precipitation may provide better constraints.

Further analyses of rolling means of rain rates and total tweets are summarized in figures 4 and 5 for Buenos Aires province in November 2014. This month is analyzed, as a flood was reported early in November. Rolling means minimize the noise and lag between tweets and rainfall events. Our results indicate a three-day rainfall event from 1-4 November corresponding with high Twitter activity. Further analyses must be performed to account for rainfall visibility and precipitation type, which may influence the perception of Twitter users' regarding how threatened they feel. This measure could reconcile the difference between relatively low rain

rates and high Twitter activity (early November), and relatively high rain rates and low Twitter activity (late November) evident in figure 4.

3.3 Seasonal analysis of Tweeting Efficiency

A metric *tweeting efficiency* is calculated to account for the inherent lull in Twitter activity during the hours most people are likely asleep. It is the ratio between average rain rate (mm/hr) and average tweet rate (tweet/hr), which results in a tweet efficiency unit of mm/tweet. We expect the highest values of tweet efficiency while people are asleep, and it is raining heavily; thus less people are awake and tweeting about rainfall. Alternatively, we expect lower values of this metric while people are generally awake and Twitter activity is high. Mathematically, rain rates should be low; however, if rain rates and tweet count have a direct relationship, with greater rain rates, the volume of tweets will also increase.

Figure 6 summarizes the hourly tweet efficiency for each province by season. Our results indicate tweet efficiency (mm/tweet) is higher before sunrise while people are likely asleep and it is raining heavily. There is no geographic trend in tweet efficiency peaks, i.e. tweet efficiency of provinces towards the east did not reach their maxima at later hours. During the winter, tweet efficiency remains consistently low throughout the day for all provinces, as it is not the warm season when convective rainfall would be expected. These results are also likely dependent upon the type of precipitation, as tweet efficiency is calculated with rain rates. Calculating tweet efficiency with volumetric rainfall may tell a different story and demonstrate different trends by province. It may also be that population density relative to the rain's location plays a role in modulating tweet efficiency, which will be examined in future work. Eventually, the ability of

tweets to indicate and discriminate severe weather pre-, during, and post-event will be quantitatively evaluated.

4. CONCLUSIONS

Subtropical South America experiences some of the most intense thunderstorms in the world, yet the physical processes governing these systems remain relatively understudied [*Zipser et al.*, 2006; *Rasmussen et al*., 2014]. Compounding to its vulnerability to flooding, this region suffers from a lack of forecasters, and sparse, incomplete data [*Rasmussen et al.*, 2013; *Choi*, 2014]. Considering a successful forecast not only involves accurate scientific knowledge, but an appropriate response by the public, it is imperative that Argentines' concerns regarding severe weather are heard. To minimize the impacts of severe weather in the region, this study leverages Twitter data to glean insight into the public's perception surrounding severe weather.

Integrating large volumes of social media and rainfall data has the potential to provide insight into behavior and perception surrounding severe weather; however, it is not a stand-alone solution. Analyses must be performed with caution and with non-invasive methods, and consider the fluidity of social norms [*Lazer et al*., 2014]. As such, specific contents in tweets were not read or analyzed. This study aimed to detect footprints of severe weather events in social media, and analyze regional tweets and quantitative information that traditional weather observations cannot detect. There is a growing body of literature evaluating the value of Twitter during natural hazards; however, these studies are limited to extraordinary events in the United States.

We analyze the frequency of each keyword queried in the Twitter data. Throughout all provinces, *rain* appears most frequently in tweets. Two terms, *wind* and *sleep*, were byproducts in the data collection, and more ubiquitous than terms included in the keyword search. Flooding was most discussed in the Autonomous City of Buenos Aires (CABA), which is likely attributed

to the dense population, urban geography and vulnerability to flooding due to poor drainage. *Rasmussen et al.* [2014] showed a large volume of media reports surrounding floods in La Pampa, but the frequency *flood* in tweets in the region is relatively low. Unlike densely populated urban cities, e.g., Buenos Aires, La Pampa is an agricultural region prone to slow-rise floods [*Latrubesse and Brea*, 2009]. The intensity and temporal differences between the two types of flooding/precipitation could influence the dialogue in the media and Twitter differently. Tweets mentioning hail were the most frequent in Mendoza, a region vulnerable to hailstorms. Previous studies have also found a maximum in media reports on hail in this province [*Rasmussen et al.*, 2014]. Interestingly, the word *alert* was infrequently tweeted in this region, with the implication that forecasts for these events are not communicated to Twitter users.

We analyze the relationship between rainfall and Twitter activity for Buenos Aires province by season. Rolling means were calculated for a $\pm$3 hour window of time to address noise in the comparison. This noise manifests itself as time offsets between the maximum rain rates and Twitter activity, as it depends whether people are tweeting more before, during, or after events. Our results show a three-day rainfall event from 1-4 November 2014 corresponding with high Twitter activity. The relationship is not linear, and factors including precipitation intensity/type, visibility/threat perception, and population density must be accounted for.

A new index, tweet efficiency, is calculated to account for the inherent lull in Twitter activity during hours while most people are likely asleep. It is the ratio between average rain rate and average tweet rate, and results in a unit of mm/tweet. We analyze tweet efficiency as a function of hour for each province. No geographic trend is detected in tweet efficiency peaks, i.e. tweet

efficiency of provinces towards the east does not reach their maxima at later hours. During the winter, tweet efficiency remains consistently low throughout the day for all provinces, as it is not the warm season when convective rainfall would be expected. These results are also likely dependent upon precipitation type. Calculating this metric with volumetric rainfall may demonstrate different trends by province.

There are several caveats and limitations to this study. While our results show evidence of a relationship between Twitter activity and rainfall, this relationship is highly dependent upon rainfall/threat visibility and population density. Furthermore, results must be analyzed with caution as we are studying human behavior, and there are likely complex and nonlinear interactions impacting the results. Even variables that can be measured definitely, e.g. rain rates, suffer biases, as satellite algorithms have been shown to perform poorly calculating deep convection over land [*Iguchi et al.*, 2009]. Studying behavior remains a challenge due to constant technological updates and the fluidity of social norms; however, our results indicate a relationship between severe weather and Twitter activity. In future studies, rainfall must be collocated with population density to ensure it is raining where people are tweeting, and precipitation/flood type may provide information about tweeters' emotions and threat perception, which are likely influencing tweeting behavior.

REFERENCES

Anabor, V., Stensrud, D.J., and de Moraes, O.L.L. (2007), Serial upstream-propagating mesoscale convective system events over southeastern South America, *Mon. Wea. Rev. 136*, 3087-3105.

Bond, R.M., Fariss, C.J., Jones, J.J., Kramer, A.D.I., Marlow, C., Settle, J.E., and Fowler, J.H. (2012), A 61-million-person experiment on social influence and political mobilization, *Nature 489, 295-298.*

Choi, S. (2014), A Report from the Field: The Impacts and Predictability of Severe Weather in Argentina, Available from: http://www.tinker.org/content/report-field-impacts-and-predictability (Accessed 25 April 2016).

Conover, M.D., Ferrara, E., Menczer, F., and Flammini, A. (2013), The digital evolution of Occupy Wall Street, *PLoS One 8.*

Davies, R. (2014), Buenos Aires Floods Update – 3 Dead and Over 5,000 Evacuated , Available from: http://floodlist.com/america/buenos-aires-floods-3-dead-5000-evacuated (Accessed 25 April 2016).

Gilbert, J. (2013), Dozens of Argentines Die in Flash Flooding, Available from: http://www.nytimes.com/2013/04/04/world/americas/record-flooding-kills-dozens-in-Argentina.html (Accessed 25 April 2016).

Golder, S.A., and Macy, M.W. (2011), Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures, *Science 333*, 1878-1881.

Guan, X., and Chen, C. (2014), Using social media data to understand and assess disasters, *Nat. Haz. 74*, 837-850.

Hagar, C. (2009), The Information and Social Needs of Cumbrian Farmers During the UK 2001 Foot and Mouth Disease Outbreak and the Role of Information and Communication Technologies, In Döring, M. & Nerlich, B. (Eds.), *The Socio-Cultural Impact of Foot and Mouth Disease in the UK in 2001: Experiences and Analyses*, Manchester University Press.

Herfort, B., de Albuquerque, J.P., Schelhorn, S.-J, and Zipf, A. (2014), Does the spatiotemporal distribution of tweets match the spatiotemporal distribution of flood phenomena? A study about the River Elbe flood in June 2013, *Proc. of the 11th International Conference on Information Systems for Crisis Response and Management (ISCRAM '14)*, University Park, PA.

Houze, R.A., Jr., Smull, B.F., and Dodge, P. (1990), Mesoscale organization of springtime rainstorms in Oklahoma, *Mon. Wea. Rev. 118*, 613-654.

Huberman, B.A., Romero, D.M., and Wu, F. (2008) Social networks that matter: Twitter under the microscope, *First Monday 14(1)*.

Hughes, A.L., and Palen, L. (2009), Twitter Adoption and Use in Mass Convergence and Emergency Events, *Proc. of the International ISCRAM Conference*, Gothenburg, Sweden.

Iguchi, T., Kozu, T., Kwiatkowski, J., Meneghini, R., Akawa, J., and Okamoto, K. (2009), Uncertainties in the rain profiling algorithm for the TRMM precipitation radar, *J. Meteorol. Soc. Jpn. 87A*, 1-30.

Java, A., Song, X., Finin, T., and Tseng, B. (2007), Why We Twitter: Understanding Microblogging Usage and Communities, *Proc. of the Knowledge Discovery and Data Mining (KDD)*, San Jose CA.

Krishnamurthy, B., Gill, P., and Arlitt, M. (2008), A Few Chirps About Twitter, *Proc. of the First Workshop on Online Social Networks*, Seattle, WA.

Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., and Cebrian, M. (2016), Rapid assessment of disaster damage using social media activity, *Sci. Adv. 2(3)*.

Latrubesse, E.M., and Brea, D. (2009), Chapter 16. Floods in Argentina, *Dev. Earth Surf. Process. 13*, 333-349.

Lazer, D., Kennedy, R., King, G., and Vespignani, A. (2014), The parable of Google Flu: Traps in Big Data analysis, *Science 343*, 1203-1205.

Leetaru, K., Wang, S., Cao, G., Padmanabhan, A., and Shook, E. (2013), Mapping the global Twitter heartbeat: the geography of Twitter, *First Monday 18(5)*.

Library of Congress (2013), Update on the Twitter archive at the library of congress. http://www.loc.gov/today/pr/2013/files/twitter_report_2013jan.pdf. Accessed 17 July 2013.

Llorente, A., Garcia-Herranz, M., Cebrian, M., and Moro, E. (2015), Social media fingerprints of unemployment, *PLoS One 10.*

Morstatter, F., Pfeffer, J., Liu, H., and Carley, K.M. (2013), Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's Firehose, *Proc. of Seventh International AAAI Conference on Weblogs and Social Media (ICWSM '13)*, Boston, MA.

Nicolini, M., and Saulo, A.C. (2006), Modeled Chaco low-level jets and related precipitation patterns during the 1997-1998 warm season, *Meteor. Atmos. Phys. 94*, 129-143.

O'Connor, B., Balasubramanyan, R., Routledge, B.R., and Smith, N.A. (2010), From tweets to polls: Linking text sentiment to public opinion time series, in *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media (ICWSM '10) Washington, DC.*

Rasmussen, K.L., and Houze, Jr., R.A. (2011), Orogenic convection in subtropical South America as seen by the TRMM satellite, *Mon. Wea. Rev. 139*, 2399-2420.

Rasmussen, K.L., Choi, S.L., Zuluaga, M.D., and Houze, R.A., Jr. (2013), TRMM precipitation bias in extreme storms in South America, *Geophys. Res. Lett. 40*, 3457-3461.

Rasmussen, K.L., Zuluaga, M.D., and Houze, Jr., R.A. (2014), Severe convection and lightning in subtropical South America, *Geophys. Res. Lett. 41 (20)*, 7359-7366.

Romatschke, U., and Houze, R.A., Jr. (2010), Extreme summer convection in South America, *J. Clim. 23*, 3761-3791.

Salio, P., Nicolini, M., and Zipser, E.J. (2007), Mesoscale convective systems over southeastern South America and their relationship with the South American low-level jet, *Mon. Wea. Rev. 135*, 1290-1309.

Simpson J., Adler R.F., and North G.R. (1988), A Proposed Tropical Rainfall Measuring Mission (TRMM) Satellite, *Bull Am Meteor Soc 69*, 278-295.

Twitter (2015), https://about.twitter.com/company (Accessed 25 April 2016).

Velasco, I., and Fritsch, J.M. (1987), Mesoscale convective complexes in the Americas, *J. Geophys. Res. 92*, 9591-9613.

Zipser, E.J, Cecil, D.J., Liu, C., Nesbitt, S.W., and Yorty, D.P. (2006), Where are the most intense thunderstorms on Earth?, *Bull. Amer. Met. Soc. 87*, 1057-1071.

Zhao, W.X., Jiang, J., Weng, J., He, J., Lim, E.P., Yan, H., and Li, X. (2011), Comparing twitter and traditional media using topic models, *Adv. in Info. Retriev. 6611*, 338-349.

FIGURES

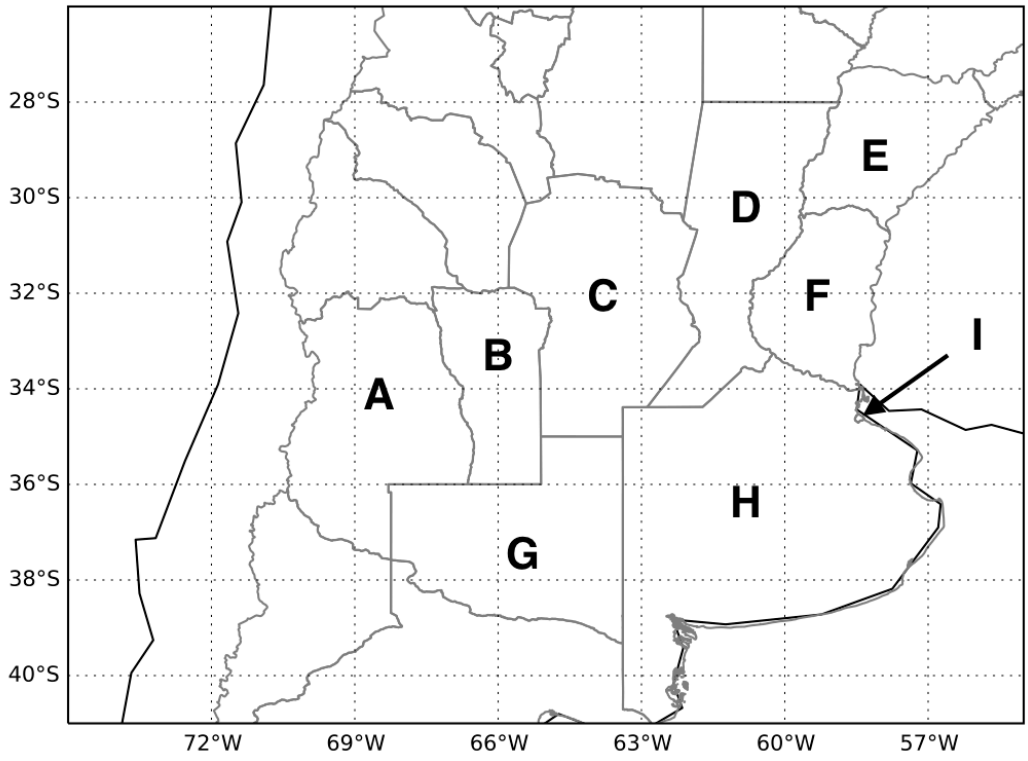

Figure 1. The nine provinces investigated for the study are: a. Mendoza; b. San Luis; c. Cordoba; d. Santa Fe; e. Corrientes; f. Entre Rios; g. La Pampa; h. Buenos Aires; i. Autonomous City of Buenos Aires.

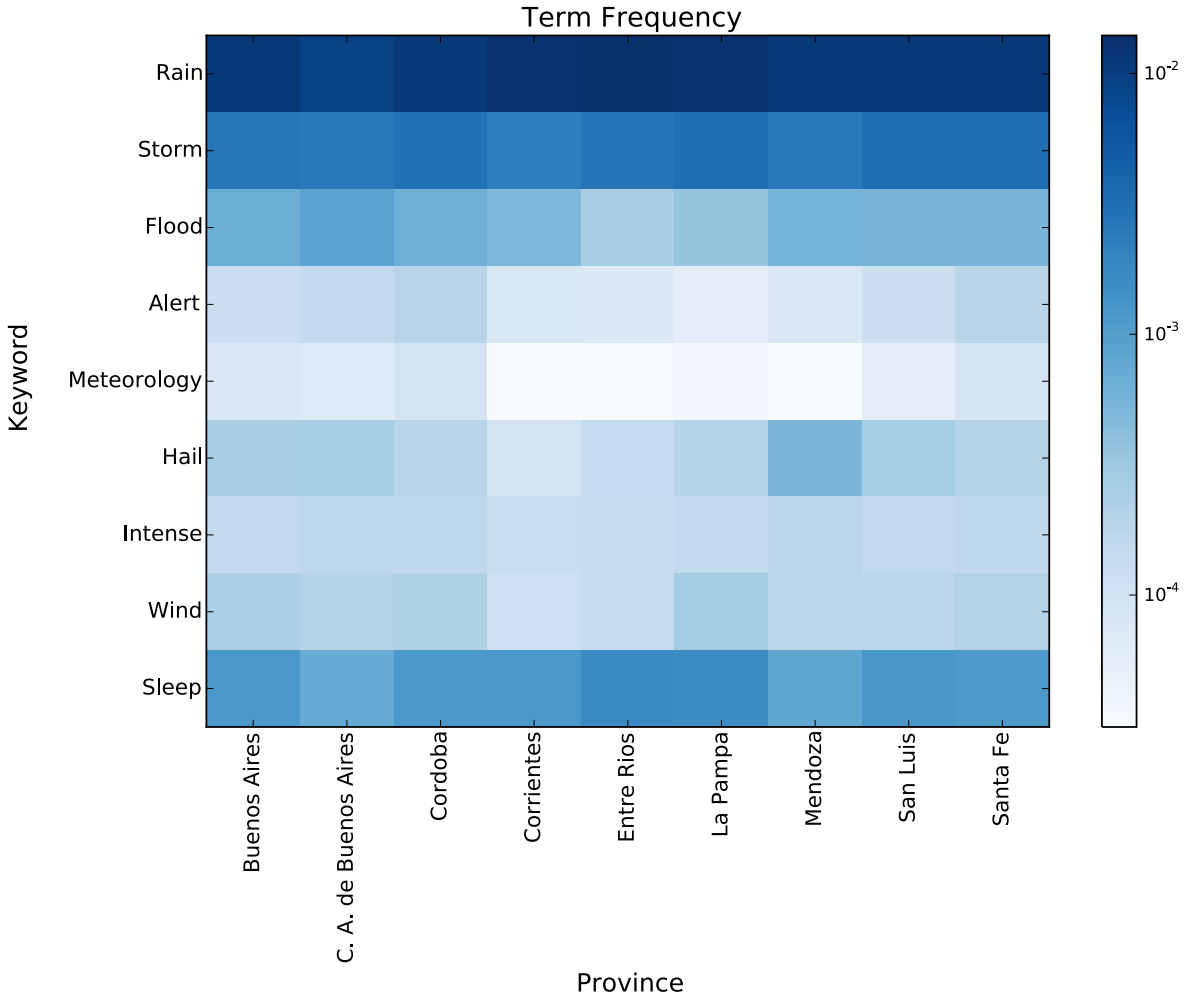

Figure 2. A heatmap of frequency of keywords mentioned in Analyzed Tweets by province.

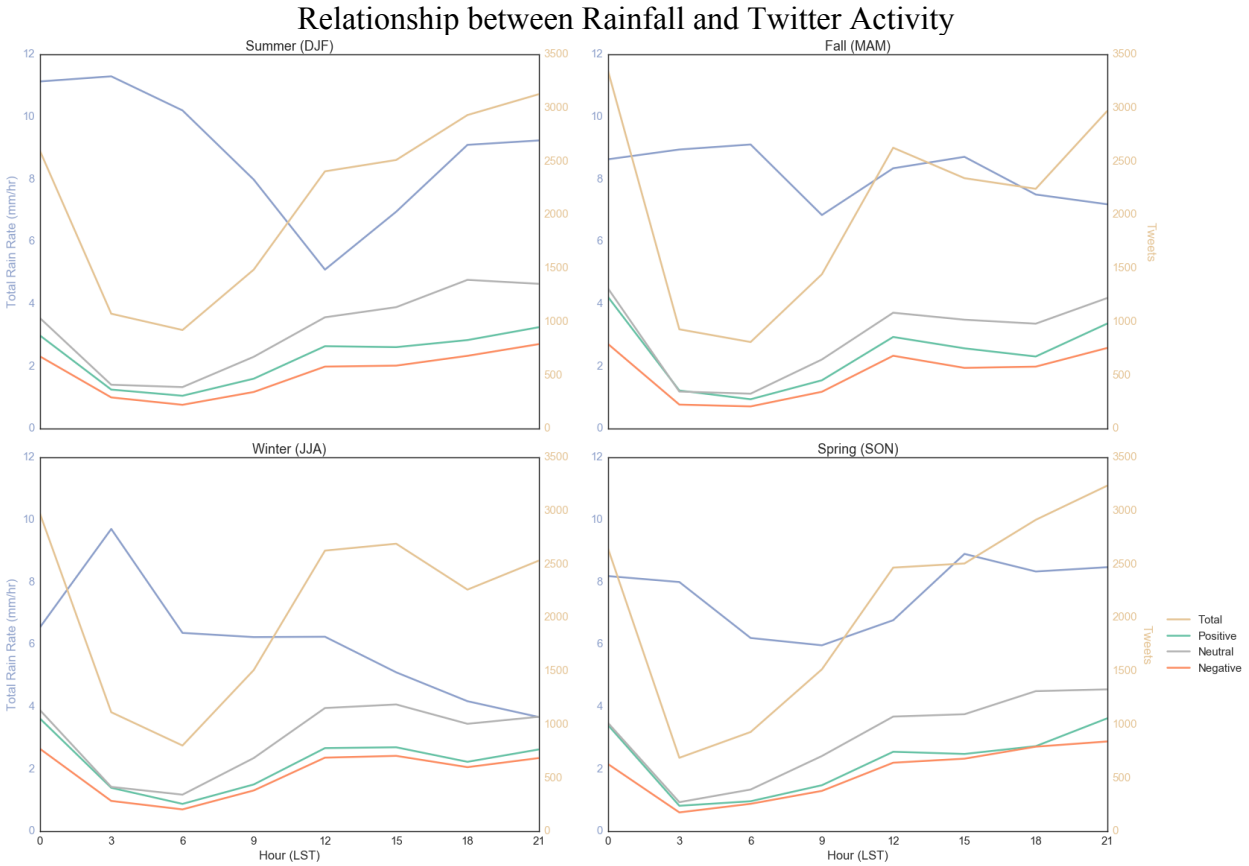

Figure 3. Hourly time series of the relationship between total rain rates and total tweets by season: DJF (summer), MAM (fall), JJA (winter), and SON (spring). The total tweet count is shown in light brown, with further analyses of tweet count by sentiment: positive, neutral and negative.

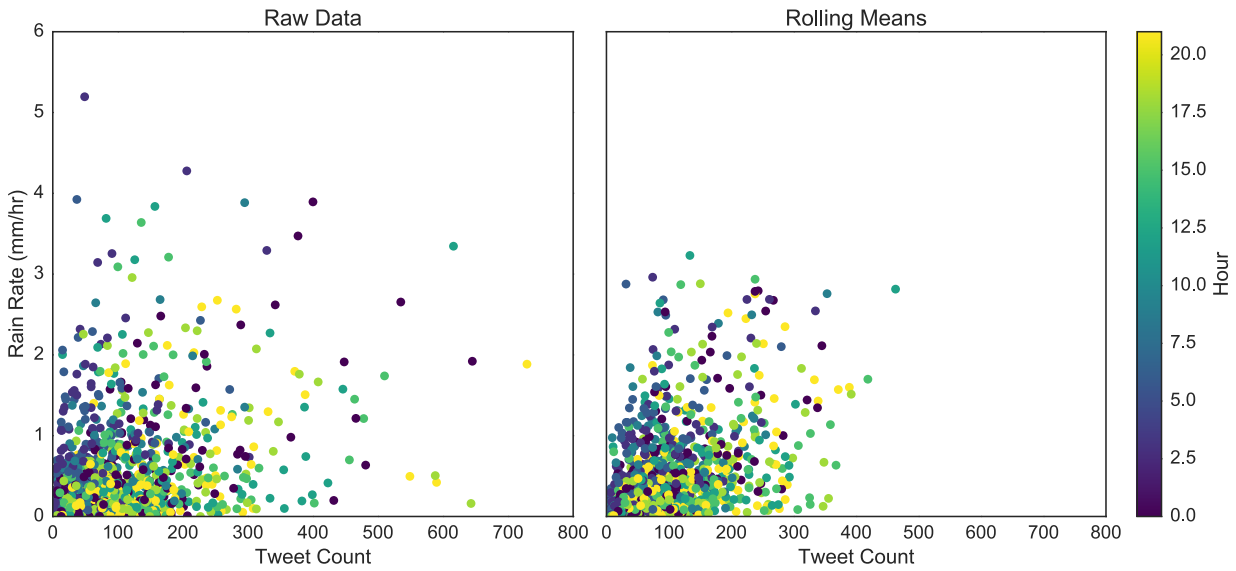

Figure 4. Average rain rates (mm/hr) as a function of tweet count in Buenos Aires province for raw data (left) and rolling means (right) by hour.

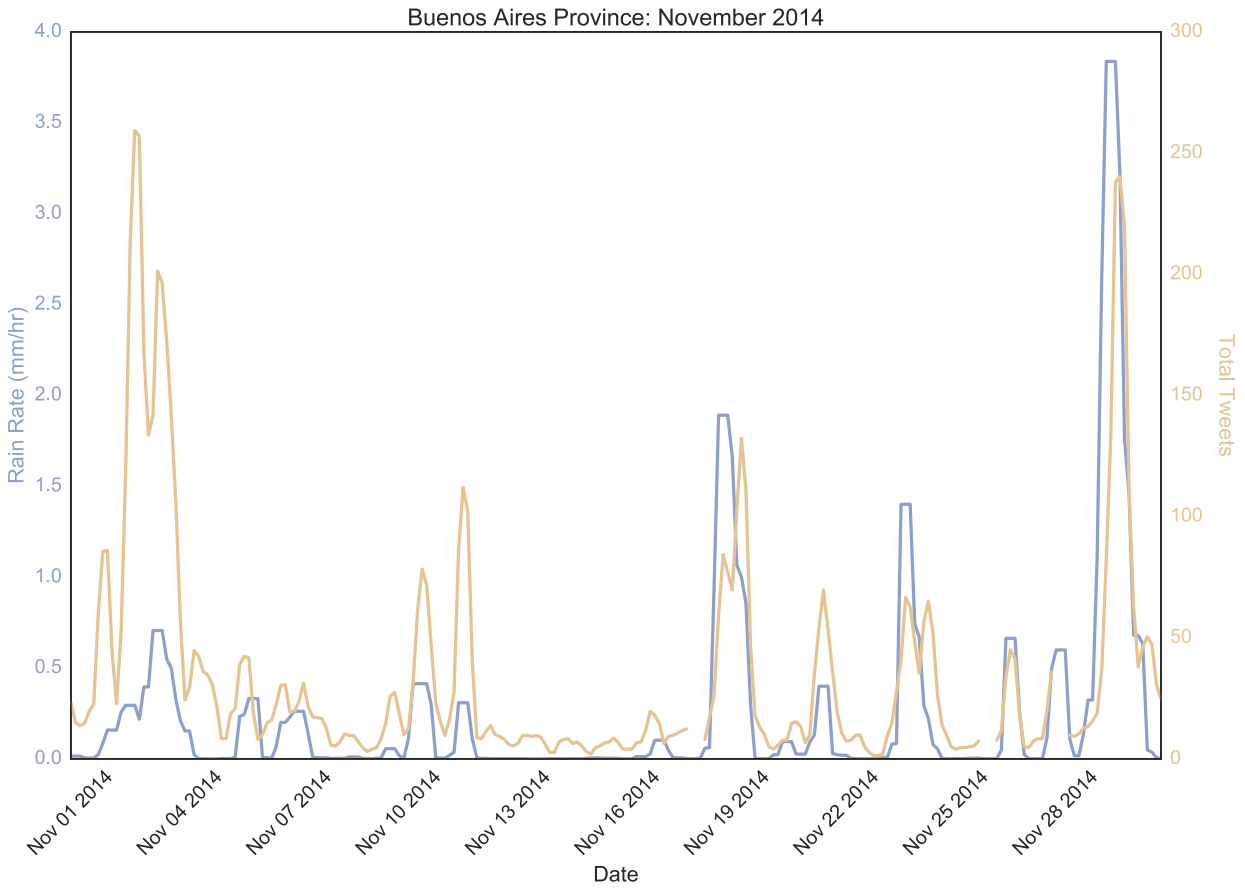

Figure 5. Rolling means of rain rate averages and total tweets for Buenos Aires Provinces during November 2014. Discontinuities indicate missing data.

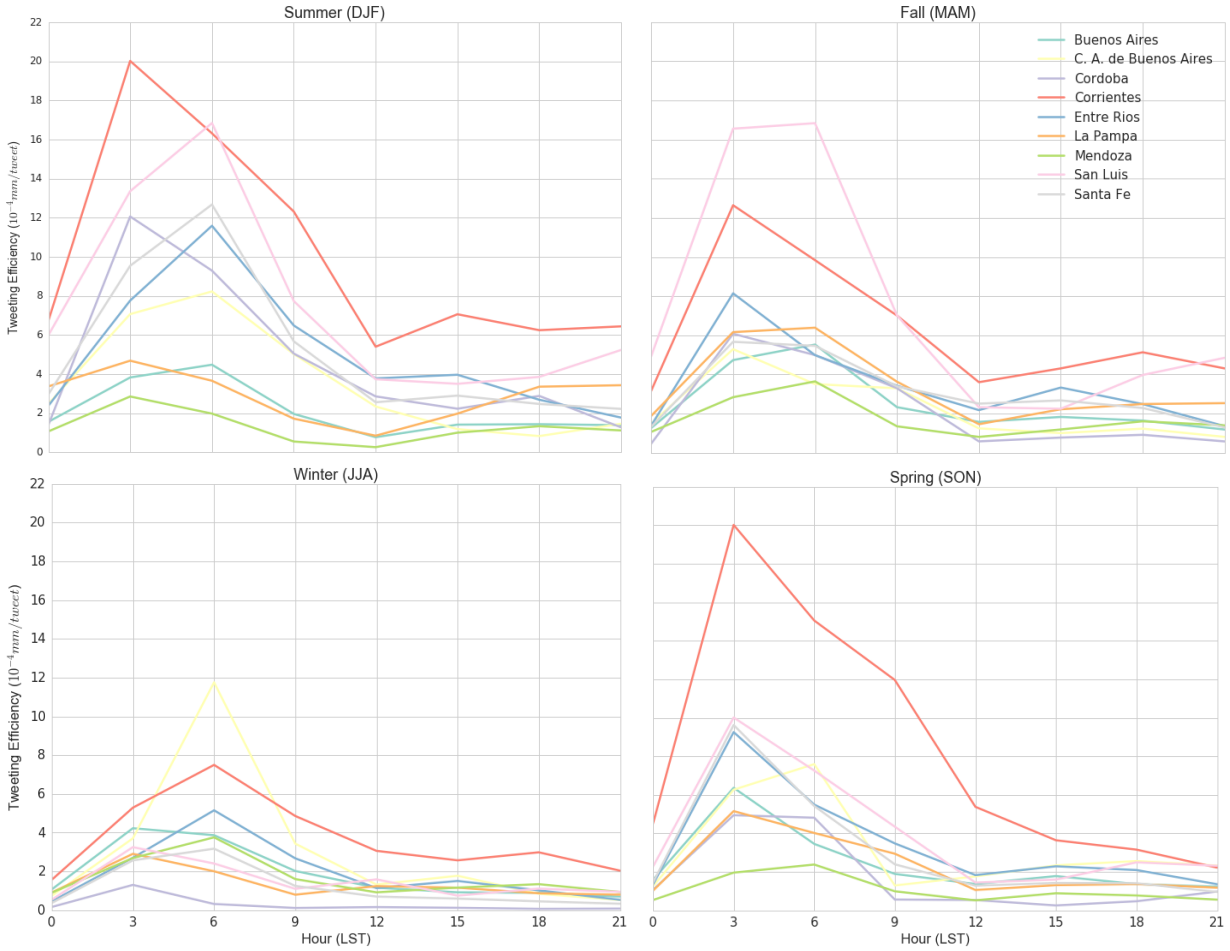

Figure 6. Hourly analysis of tweeting efficiency for all provinces by season.

TABLES

Province	Total Filtered Tweets	Analyzed Tweets	Percent (%)
Autonomous City of Buenos Aires	6450499	230024	3.57
Buenos Aires	1797588	227770	12.67
Cordoba	897631	227082	25.30
Corrientes	154772	134818	87.11
Entre Rios	342907	210923	61.51
La Pampa	127888	118847	92.93
Mendoza	207416	159630	76.96
San Luis	71268	65713	92.21
Santa Fe	875760	227079	25.93

Table 1. Total Filtered Tweets found, Analyzed Tweets after posts by bots were removed, and percentage of Analyzed Tweets to Total Filtered Tweets that met criteria by province.

Province	Northern Latitude (°)	Southern Latitude (°)	Western Longitude (°)	Eastern Longitude (°)
Autonomous City of Buenos Aires	-34.5	-34.7	-58.5	-58.3
Buenos Aires	-33.9	-40.9	-63.4	-56.9
Cordoba	-29.68	-35.0	-65.5	-62.2
Corrientes	-27.4	-30.5	-59.6	-55.6
Entre Rios	-30.4	-33.9	-60.4	-57.7
La Pampa	-35.0	-38.9	-68.3	-63.4
Mendoza	-32.1	-37.3	-70.2	-67.0
San Luis	-31.9	-35.9	-66.6	-65.0
Santa Fe	-27.9	-34.4	-62.9	-58.9

Table 2. Latitude and longitude values for bounding box of each province used retrieve rain rates.

Province	R^2
Autonomous City of Buenos Aires	0.1710
Buenos Aires	0.4256
Cordoba	0.3514
Corrientes	0.3645
Entre Rios	0.5064
La Pampa	0.3659
Mendoza	0.2454
San Luis	0.3082
Santa Fe	0.3916

Table 3. R-squared values for the rolling averages of tweet count and rain rates.